Statistical Computation Algorithm Based on Markov Chain Monte Carlo Sampling to Solve Multivariable Nonlinear Optimization

Tengzhong Rong *1, Chaolin Liu2

College of Mathematics and Statistics, Chongqing University, Chongqing, China

1timepdf@gmail.com; 2lcl@cqu.edu.cn

***Abstract*-This paper proposes a method to solve multivariable nonlinear optimization problem based on MCMC statistical sampling. Theoretical analysis proves that the maximum statistic converges to the maximum point of probability density. From this convergence property, we establish a newly algorithm to find the best solution of an optimization through largely MCMC sampling in condition that the exact generalized probability density function has been designed. The MCMC optimization algorithm needs less iterate variables reserved so that the computing speed is relatively high. Finally, we build a Markov chain model to demonstrate the dynamic movement of ladderlike labour transfer with multiparamter, and we practice this MCMC sampling optimization algorithm to find parameters estimations. The optimization results are reasonable and from which we obtain the short-term and long-term forecast of labour movements.**

Keywords-Multivariate Nonlinear Optimization; MCMC Sampling; Statistical Computation; Generalized Function; Labour Transfer

I. INTRODUCTION

Statistical computing is an active frontier in modern statistics which focus on the application of calculation technology in statistics. In many complex stochastic problems, the theoretical probability is difficult to obtain directly because of the complexity of random factors. The solution is often replaced by the corresponding probability sample frequency analysis from computer simulation, which is known as the Monte Carlo method. The conditional distribution method is based on Markov chain samples generated indirectly when confront sampling from the complex distribution. It generates random samples whose are not independent, the new sample is correlated with previous one. It can be shown that on the ergodic condition, the limiting distribution of this Markov chain is exactly the distribution we intend to sample. This method is known as Markov Chain Monte Carlo (MCMC) algorithm.

Metropolis-Hasting[1] algorithm is the first application of a MCMC sampling algorithm which is mainly used for simulation and sampling of single dimensional random variable. Gibbs sampling[2] is also a common and representative MCMC algorithm which is mainly used for multi-dimensional random variables sampling.

MCMC algorithms have been widely used in statistical inference for many areas. In 1998, Brooks[3] discussed the continuous and discrete variables MCMC sampling techniques systematically, and he found answer for the iteration problem of stationary distribution when the MCMC algorithm is expanded for the Baysian parameter estimation and figure computing applications. In 2003, Andrieu[4] made a further study on machine learning and parameter inference with this random sampling method. MCMC algorithm also has important applications in the measurement of management such as time series model selection[5], Bayesian statistical inference[6], measuring the model parameter estimation[7] etc. Currently when we apply MCMC algorithm to estimate parameter, the common method is frequency consistency by law of large numbers and estimate the distribution expectation by the sample average, but for other sample statistics such as maximum, median, etc. , there are few application.

Multivariable nonlinear optimization problem is hard to find the best solution. The traditional methods include the genetic algorithm[8], the ant colony algorithm[9]，simulated annealing[10] and so on. Only a few works involved with MCMC algorithm in optimization problem. In 1995, Eyer[11] applied MCMC sampling to simulated annealing search design. He simulated annealing random variation of process and control chain convergence to optimal solutions. But Eyer's[11] method still focuses on the purpose of convergence, it accept poor disturbance in certain probability and take the final convergence result as the best solution. Because the selection of parameter annealing has significantly impact on the solution, and the rule of "probability accept" may results in the current state less than some solutions among the search trajectory, It is difficult to obtain global optimization solution. Even in some circumstance, the final algorithm approximate optimal solution is worse than the solution in search trajectory.

The rest of this paper will be organized as follows: In section II, we will introduce the tradition methods of MCMC sampling, that is, Metropolis-Hasting Sampling Algorithm and Gibbs Sampling Algorithm. In section III, we will propose a computation statistics algorithm for multivariable and nonlinear optimization problem by the large sample properties of MCMC statistical method. This newly thinking for MCMC application includes the design of generalized density function, using of Metropolis-Hasting and Gibbs sampling methods, calculation the maximum statistic. At last, we will show the application of our MCMC Sampling algorithm to solve the real problem in section IV.

II. MCMC SAMPLING ALGORITHM

A. *Metropolis-Hasting Sampling Algorithm*

The basic idea of MCMC methods is to construct a Markov chain with the specified stationary distribution, namely $\pi(x)$, then run the chain with full length till the sample chain value close enough to its stationary distribution. Then take stationary chains as the samples of $\pi(x)$ and

make variety of statistical inference based on these samples. The first MCMC sampling method is Metropolis-Hasting[1] algorithm, which means sampling starts from another easily known reversible Markov chain Q, and obtain the new Markov chain by comparing. Suppose $q(i, j)$ indicated that the transfer from the state i to j, and note x_0 the initial point, then this algorithm is as follows:

1) Drawn y from the known distribution $q(x,y)$;
2) Drawn u from uniform distribution $U(0,1)$
3) Compute $r(x_t, y) = \min\left\{1, \dfrac{\pi(y)q(y, x_t)}{\pi(x_t)q(x_t, y)}\right\}$
4) $x_{t+1} = \begin{cases} y & u \le r \\ x_t & u > r \end{cases}$

Repeat 1) to 4), Hastions[1] had proved that the transfer accordance with 1) to 4) is reversible Markov chain has stationary distribution π（x）. In particular of convenience, the transition probability $q(i, j)$ can be a special conditional independence case, that is, the transition probability only relevant with j. Here $r(x_i, y)$ can be simplified as $\min\left\{1, \dfrac{\pi(y)q(y)}{\pi(x_t)q(x_t)}\right\}$.

B. Gibbs Sampling Algorithm

Gibbs sampling algorithm was first proposed by Geman[2] in 1984, it is a special case of Metropolis-Hastings sampling. Gibbs sampling method is to separate the vector space into multiple parts and each part of the transformation to move forward by sampling. Gibbs sampling algorithm is suitable for high-dimensional random variables especially.

The Gibbs algorithm is as follows: Assumed $X=(X_1,X_2,...X_k)$ to be k-dimensional random variables with joint distribution π, π_i $(\pi_i\ (X_i/X_1,..,X_{i-1},X_{i+1}..,X_k)$ is the conditional distribution of X_i. Gibbs sampling is to generate random samples X_i according to the distribution π_i with condition $(X_1^{t-1}, X_2^{t-1}, \ldots, X_k^{t-1})$, where

$$X_t^i \sim \pi_i(X_i \mid X_1^t, \ldots, X_{i-1}^t, X_{i+1}^{t-1} \ldots, X_k^{t-1}),\ i = 1, 2, \ldots, k$$

Then we obtain $(X_1^{t-1}, X_2^{t-1}, \ldots, X_k^{t-1})$, and that is the whole process of Gibbs sampling. Gibbs sampling method converts high-dimensional distribution sampling into one-dimensional conditional distribution sampling. With iteration, the new high-dimensional samples $(X_1^{t-1}, X_2^{t-1}, \ldots, X_k^{t-1})$ are random sampling from distribution π, which avoid the difficult to sample high dimensional distribution directly.

III. MCMC STATISTICAL COMPUTING OPTIMIZATION ALGORITHM

In a probability distribution, the largest density area is mostly tending to be sampled. So the sampling density function should converge to the maximum point of maximum probability if the sample is sufficiently large. That establishes the link between the function maximum value and sampling extreme statistics.

A. Maximum Statistics of One-Dimensional Density

Note x_0 a maximum point of a continuous density $f(x)$, $x \in R$, that is $f(x_0) = \max\limits_{x \in R} f(x)$. With the idea of likelihood, $(x_1, \cdots, x_n)$, the n-times the sample from $f(x)$, take x^*, which is $f(x^*) = \max\left(f(x_1), \cdots, f(x_n)\right)$,as estimation of x_0. Make $F_X(x)$ the distribution function of $f(x)$, $X_1, \cdots, X_n$ is simple sample and note $Y = f(X)$, that $Y_i = f(X_i), i = 1, 2, \cdots, n$ are independent and identically distribution. We have

$$F_Y(y) = P\{f(X) < y\} = \int\limits_{f(x) < y} f(x)dx \quad 0 \le y \le f(x_0) \quad (1)$$

By the continuous density $f(x)$, $F_Y(y)$ is increasing strictly on $0 \le y \le f(x_0)$, and $F_Y(0) = 0$, $F_Y(f(x_0)) = 1$. Take $Y_n^* = f(X^*) = \max\left(f(X_1), \cdots, f(X_n)\right)$, we have

$$F_{Y_n^*}(y) = \left[F_Y(y)\right]^n, \quad 0 \le y \le f(x_0) \quad (2)$$

So when $n \to \infty$ and $0 \le y < f(x_0)$, then $F_{Y_n^*}(y) \to 0$; when $y = f(x_0)$, then $F_{Y_n^*}(y) \to 1$. That is, Y_n^* convergence to $f(x_0)$ in probability and estimator x^* convergence x_0 in probability equivalently.

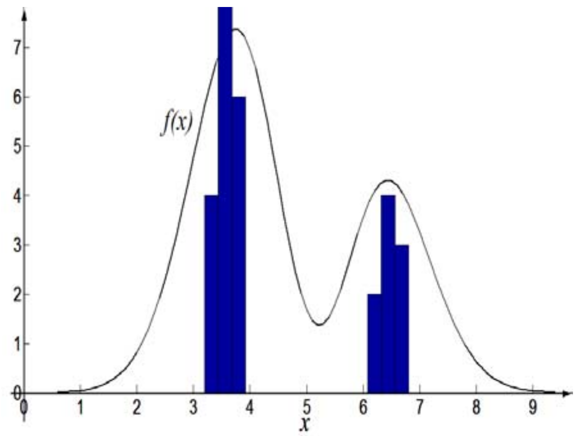

Fig. 1 Sampling from probability density function

B. Maximum Statistics of Multi-Dimensional Density

Set x_0 a maximum point of a continuous density $f(x)$, $x \in R^n$. The n-times the sample $(x_1, \cdots, x_n)$ from $f(x)$, take x^*, which is $f(x^*) = \max\left(f(x_1), \cdots, f(x_n)\right)$, as estimation of x_0. Note $Y = f(X)$, then

$$F_Y(y) = \int\limits_{f(x)<y}\!\!\ldots\!\int f(x)dx, \quad 0 \le y \le f(x_0) \quad (3)$$

By the continuous density $f(x)$, $F_Y(y)$ is increasing strictly on $0 \le y \le f(x_0)$, and $F_Y(0) = 0$, $F_Y(f(x_0)) = 1$. Take $Y_n^* = f(X^*) = \max\left(f(X_1), \cdots, f(X_n)\right)$, we have

$$F_{Y_n^*}(y) = \left[F_Y(y)\right]^n, \quad 0 \le y \le f(x_0) \quad (4)$$

Therefore when $n \to \infty$ and $0 \le y < f(x_0)$, then $F_{Y_n^*}(y) \to 0$; when $y = f(x_0)$, then $F_{Y_n^*}(y) \to 1$. That is, Y_n^* convergence to $f(x_0)$ in probability when $n \to \infty$ and estimator x^* convergence x_0 in probability equivalently.

Similar conclusions can be deduced to one-dimensional and multi-dimensional discrete distribution.

C. Generalized Probability Density Function Design

Suppose the variables space of an optimization problem is n-dimensional feasible region $x \in \Theta$, and the objective function is $\max g(x)$.

(1) If $g(x) \geq 0, x \in \Theta$, note the generalized probability density function as

$$\pi(x) = \frac{g(x)}{\int\limits_{x \in \Theta} g(x)dx}, \quad x \in \Theta \tag{5}$$

$\pi(x), x \in \Theta$ can be considered as a generalized probability density function and its maximum point corresponds to the maximum point of $g(x)$. Thus through the previous analysis, x^* will continue to close and converge to the maximum point of $\pi(x)$ in distribution with the larger sampling. That is MCMC statistical computing optimization algorithm to solve multivariate objective function optimization problem.

(2) If $g(x) \geq 0, x \in \Theta$ is not satisfied, the generalized probability density function can be transformed as

$$\pi(x) = \frac{e^{g(x)}}{\int\limits_{x \in \Theta} e^{g(x)}dx}, \quad x \in \Theta \tag{6}$$

(3) If the objective function is $\min g(x)$, the generalized probability density function can be transformed as

$$\pi(x) = \frac{e^{-g(x)}}{\int\limits_{x \in \Theta} e^{-g(x)}dx}, \quad x \in \Theta \tag{7}$$

D. MCMC Optimization Algorithm Design

Now the difficulty is that multivariate distribution $\pi(x), x \in \Theta$ is relatively complicated and it is not easy to practice sampling. While multi-dimensional distribution of random sampling with MCMC method has better adaptability, the sampling problem can be easily removed.

Suppose we have the probability density function (5), the other two have almost the same induction as follows. Design a Markov chain with stationary distribution is $\pi(x), x \in \Theta$, then the maximum point in finite sampling from distribution $\pi(x)$ will be sufficiently close to the maximum point of $g(x)$ in the feasible region, in which state $x \in \Theta$ of the Markov chain is n-dimensional vector.

First let $T(x)$ is the uniform distribution on $x \in \Theta$, and sampling from $T(x)$ is through Gibbs sampling. Because there have many restrictions on feasible region, the region $x \in \Theta$ is an irregular n-dimensional area and its samples can be generated through Gibbs conditional distribution. Then, since $T(x_t) = T(y)$, by the MCMC algorithm we have

$$r(x_t, y) = \min\left\{1, \frac{\pi(y)T(x_t)}{\pi(x_t)T(y)}\right\} = \min\left\{1, \frac{g(y)}{g(x_t)}\right\} \tag{8}$$

The optimization algorithm is:

1) Drawn y by the Gibbs method from the distribution $T(x)$;

2) Drawn u from the uniform distribution $U(0,1)$

3) Compute $r(x_t, y) = \min\left\{1, \frac{g(y)}{g(x_t)}\right\}$

4) Take $x_{t+1} = \begin{cases} y & u \leq r \\ x_t & u > r \end{cases}$

5) Take $x_{t+1}^* = \begin{cases} x_t^* & g(x_t^*) > g(x_{t+1}) \\ x_{t+1} & g(x_t^*) \leq g(x_{t+1}) \end{cases}$

Repeat 1) to 5). If the iteration times are large enough, then x_n^* will convergence to the maximum point of the objective function $g(x)$ in distribution.

We can see from the problem analysis above that the key points of MCMC statistical computing method are designing of general probability density function and uniform sampling from conditional constraint region.

IV. NUMERICAL EXAMPLE

In current western china, the surplus labour force transfer becomes the key point to achieve the balance between urban and rural. Several researchers have theoretically investigated the surplus labour force transfer from other perspective. Taylor and Williamson[12] investigated the labour force transfer from the perspective of population age structure in 1997, and in 2005, Temple[13] from the perspective of agricultural economic development. Fields [14] considered the labour immigrant for labour remuneration compensation, and Carrasco *et al.*[15] in 2008 studied the labour market effect of immigrant for employment rates and wages. All of these studies are qualitative but not quantitative analysis especially for prediction. In the following we will propose a Markov chain to express the rural labour transfer stochastically.

A. Model of Ladderlike Labour Transfer

Ladderlike labour transfer mode is developed in the work of balance urban and rural in Chongqing, a young municipality in western china. It means in a county region, the rural labour in remote countryside free themselves from traditional farming operations and transfer to new rural residential point, small town, and city to work in secondary and tertiary industries gradually.

By considering labour as the transfer variable, different terrain shift of labour employment region, such as countryside, residential point, small town, city as the variable state, we build a labour transfer model. In view of the labour in each state could transfer other places out side this county region, we propose an additional variable to present it. We denote the five states as 1,2,3,4 and 5 individually. From Fig. 3 we can see the transfer movements between each state. Labour may stay at the same state or transfer to lower ladder state. State 1,2,3,4 may turn to 5 but state 5 can only turn back to 4 because those guys work in huge cities unwilling to go back rural again. The ladderlike transfer process is not always in

the same direction, but there have some reciprocating mobility to be a complex dynamic process.

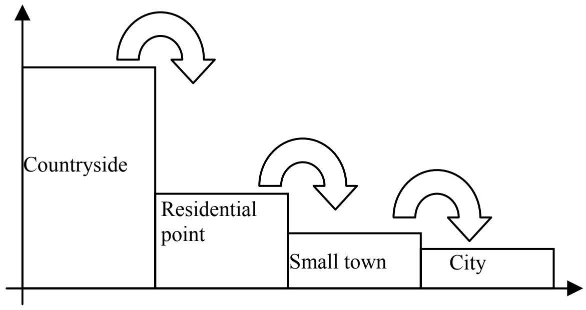

Fig. 2 The different ladders of the labour transfer

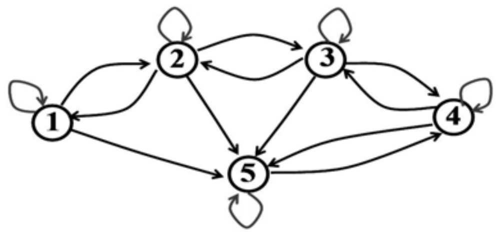

Fig. 3 The state transition diagram of labour transfer

Thus a discrete homogeneous Markov chain is built, and the labour transfer is determined by the transition probability matrix $P=(p_{ij})$.

$$P=\begin{bmatrix} p_{11} & p_{12} & & & p_{15} \\ p_{21} & p_{22} & p_{23} & & p_{25} \\ & p_{32} & p_{33} & p_{34} & p_{35} \\ & & p_{43} & p_{44} & p_{45} \\ & & & p_{54} & p_{55} \end{bmatrix} \tag{9}$$

B. MCMC to Estimate the Transition Probability Matrix

Let $\Pi_i=(x_{i1},x_{i2},\cdots,x_{is}),\quad i=1,2,\cdots,n$ be the n steps actual transfer observation samples, by observational error ε, there have

$$\Pi_{i+1}=\Pi_i P+\varepsilon_i,\quad i=1,2,\cdots,n-1 \tag{10}$$

To estimated transition probability matrix P is a continuity optimization problem. Although estimation P can be regarded as a linear model, but P has a restriction that summation of each row is 1. Therefore, the optimization cannot be solved by traditional method. The parameter solving can be regarded as 16 variables optimization problem, namely $P=(p_{ij})$ should satisfy

$$\min \sum_{i=1}^{n-1}(\Pi_{i+1}-\Pi_i P)(\Pi_{i+1}-\Pi_i P)^T$$
$$s.t.\quad \begin{cases} p_{ij}\geq 0 & i,j\in S \\ \sum_{j\in S} p_{ij}=1 & i\in S \end{cases} \tag{11}$$

On the basis of the MCMC algorithm design, the first step is application Gibbs sampling multivariate $(p_{11},p_{12},\cdots,p_{1s},p_{21,}\cdots,p_{2s},p_{s1}\cdots p_{ss})$ with the condition $P=(p_{ij})$ by uniform sampling. Design sampling method as: start from $(T_{11},T_{12},\cdots,T_{1s},T_{21,}\cdots,T_{2s},T_{s1}\cdots T_{ss})$, which $T_{ij}\sim U(0,1)$. And normalize it $p_{ij}=\frac{T_{ij}}{\sum_{j\in S}T_{ij}}$, so $(p_{11},p_{12},\cdots,p_{1s},p_{21,}\cdots,p_{2s},p_{s1}\cdots p_{ss})$ was a qualified distribution. By the Gibbs sampling method, sample $T_{ij}\sim U(0,1)$ from $(T_{11},T_{12},\cdots,T_{1s},T_{21,}\cdots,T_{2s},T_{s1}\cdots T_{ss})$ and normalized successively. Then obtain a new eligible $(p_{11},p_{12},\cdots,p_{1s},p_{21,}\cdots,p_{2s},p_{s1}\cdots p_{ss})$ sample. At last, sample largely by MCMC algorithm and record the maximum statistic. From the theoretical analysis we know it will converges to the optimum of the objective function.

We use the original labour distribution sample $\Pi_i=(x_{i1},x_{i2},\cdots,x_{is}),\quad i=1,2,\cdots,n$ from 2001 to 2011, sample iterate from the generalized density function using MCMC statistical optimization algorithm to obtain optimized parameters estimation $\hat{P}$.

$$P=\begin{bmatrix} 0.782 & 0.079 & & & 0.138 \\ 0.051 & 0.457 & 0.295 & & 0.196 \\ & 0.108 & 0.417 & 0.316 & 0.158 \\ & & 0.244 & 0.610 & 0.146 \\ & & & 0.549 & 0.451 \end{bmatrix} \tag{12}$$

From the estimation $\hat{P}$ we can see that $\hat{p}_{11}$ is 0.782, which indicate the labour in countryside remain unchanged at the probability 0.782, and $\hat{p}_{15}$ equals to 0.138 means the probability of go out of the county to work is 0.338. From $\hat{p}_{23}$ equals to 0.295 we can see labour in resident transfer to small town at probability 0.295, which means these labours have adapt to urban life and accelerated the transfer consciously. $\hat{p}_{21}$ equals to 0.051 indicate there still have one percent rural labour return back to countryside because they are not suited the life in residential point. About 0.601 parts of labour in city remain unchanged, which includes the regular citizen. Probability of $\hat{p}_{35}$ is 0.158 represent those who work and settle down outside the county. We can see the estimated parameters in transition probability matrix have characteristic of ladderlike transfer, so the MCMC estimation is reasonable.

C. Prediction of Labour Transfer

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

From (10) we count Π iteratively 3 times to obtain the short term prediction of labour distribution in the next 3 years.

TABLE I
THE SHORT TERM PREDICTION OF LABOUR PROBABILITY DISTRIBUTION IN THE NEXT 3 YEARS

Year	Countryside	Residential point	Small town	City	Outside county
2012	0.406	0.153	0.182	0.237	0.221
2013	0.325	0.122	0.179	0.323	0.249
2014	0.261	0.101	0.190	0.391	0.257

We can see from Table I that the rural labour will transfer to small town and cities continuously in short period, the proportion of labour in countryside will gradually decrease

but the decrease rate will slow down. The labour proportion in outside the county keeps increasing and the proportion in residential point fluctuates slightly.

We also obtain the stationary distribution of this Markov chain as the long term prediction of each state, which is 0.0125, 0.0476, 0.2302, 0.4926 and 0.2174. We can see in current policy and economic circumstance, the labour in city are in the majority, there only leave about 0.0125 proportion labour stay in countryside in the far future. That shows if the local government keep the current economic and policy unchanged for a long time, city-countryside dualization can eliminated.

V. CONCLUSIONS

The algorithm based on MCMC sampling method of statistical optimization is demonstrated to solve multi-variable nonlinear optimization problem. The key points of the MCMC optimization algorithm are the generalized probability density function design and Gibbs sampling algorithm from feasible region. MCMC optimization algorithm does not exist local optimal solution problem, and theoretically ensures the global optimal solution can be obtained if sampling large enough. We practice the MCMC optimization method to deal with parameter estimations for labour transfer model. Those estimations seem reasonable and from which we achieve a short-term and long-term forecast for labour circumstance.

ACKNOWLEDGMENT

We wish to thank the anonymous reviewers for their helpful comments. This research was supported by Chongqing University Teaching Reform Project under Grand No. 201104HS.

REFERENCES

[1] W. K. Hastions, “Monte carlo sampling methods using Markov chains and their applications,” Biometrika, vol.57, 1970, pp. 97-109.

[2] S. Geman, and D. Geman, “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, no. 6, 1984, pp. 721-741.

[3] S. P. Brooks, “Markov chain monte carlo method and its application,” Journal of the Royal Statistical Society. Series D (The Statistician), vol. 47, no. 1, 1998, pp. 69-100.

[4] Andrieu, et al., “An introduction to MCMC for machine learning,” Machine Learning, vol. 50, no. 1-2, 2003, pp. 5-43.

[5] X. Zhao, “ARMA model selection based on monte carlo markov chain,” Application of Statistics and Management, vol. 02, 2006, pp. 161-165.

[6] C. J. Perez, J. Martin, and M. J. Rufo, “Sensitivity estimations for Bayesian inference models solved by MCMC methods,” Reliability Engineering & System Safety, vol. 91, no. 10, 2006, pp. 1310-1314.

[7] H. Wago, “Bayesian estimation of smooth transition GARCH model using Gibbs sampling,” Mathematics and Computers in Simulation, vol. 64, no. 1, 2004, pp. 63-78.

[8] C. R. Duke, et al., “Optimization of control source locations in free-field active noise control using a genetic algorithm,” Noise Control Engineering Journal, vol. 57, no. 3, 2009, pp. 221-231.

[9] Aghaie, and H. Mokhtari, “Ant colony optimization algorithm for stochastic project crashing problem in PERT networks using MC simulation,” International Journal of Advanced Manufacturing Technology, vol. 45, no. 11-12, 2009, pp. 1051-1067.

[10] M. Yogeswaran, S. G. Ponnambalam, and M. K. Tiwari, “An efficient hybrid evolutionary heuristic using genetic algorithm and simulated annealing algorithm to solve machine loading problem in FMS,” International Journal of Production Research, vol. 47, no. 19, 2009, pp. 5421-5448.

[11] G. Eyer, et al. “Annealing Markov chain monte carlo with applications to ancestral inference,” Journal of the American Statistical Association, vol. 90, no. 431, 1995, pp. 909-920.

[12] Taylor, and J. Williamson, “Convergence in the age of mass migration,” European Review of Economic History, vol. 1, 1997, pp. 27-63.

[13] J. Temple , “Growth and wage inequality in a dual economy,” Bulletin of Economic Research, vol. 57, no. 2, 2005, pp. 145-169.

[14] G. S. Fields, “A wellfare economic analysis of labor market policies in the Harris-Todaro Model,” Journal of Development Economics, vol. 76, no. 1, 2005, pp. 127-146.

[15] R. Carrasco, J. F. Jimeno, and A. C. Ortega, “The effect of immigration on the labor market performance of native-born workers: some evidence for Spain,” Journal of Population Economics, vol. 21, no. 3, 2008, pp. 627-648.